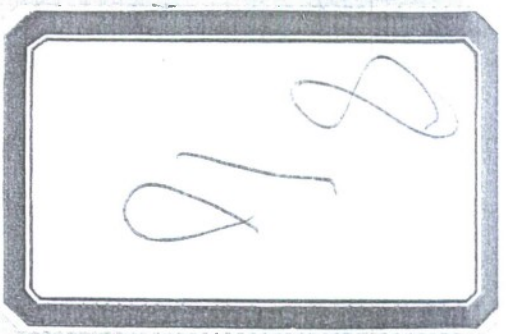

DEVELOPMENT AND VALIDATION OF A CHECKLIST FOR MEASURING SUBJECTIVE FATIGUE

56-115

Air University
SCHOOL OF AVIATION MEDICINE, USAF
RANDOLPH AFB, TEXAS

THE DEVELOPMENT AND VALIDATION OF A CHECKLIST FOR MEASURING SUBJECTIVE FATIGUE

RICHARD G. PEARSON, First Lieutenant, USAF
GEORGE E. BYARS, JR., Airman First Class
Department of Psychology

56-115

Air University
SCHOOL OF AVIATION MEDICINE, USAF
RANDOLPH AFB, TEXAS
December 1956

THE DEVELOPMENT AND VALIDATION OF A CHECKLIST FOR MEASURING SUBJECTIVE FATIGUE

INTRODUCTION

The problem of aircrew fatigue is becoming more acute as today's flights become longer and more complex. Yet, as important as this problem is, no reliable measure of fatigue has been developed. Although the physiologic and performance decrement aspects of fatigue have been the subject of numerous studies in physiology and psychology (1, 2, 3), the many difficulties which have turned up in dealing with the so-called "objective" measures of fatigue call for a renewed attack upon a third aspect of fatigue, the subjective response (4-8).

The possibility of establishing a subjective measure of fatigue has seemed unpromising because of the difficulty of treating affective responses in a simple, quantitative manner. Although several studies employing measures of subjective fatigue have been reported in the literature (8-14) it has only been within the last few years that any study has been designed primarily to develop a fatigue scale. Frazier (15) attempted to scale 70 items by Guttman's method (16) but was unsuccessful. McNelly (17) added to Frazier's list, then scaled 123 items by Thurstone's method of equal-appearing intervals. Two 9-item equivalent-form scales were then constructed and validated in a laboratory study. The research to be described in this paper, however, was initiated prior to the availability of McNelly's results, and was based on a somewhat different definition of the fatigue continuum.

If a scale of subjective fatigue is to be meaningful its items should fall along a single dimension; that is, the scale should be unidimensional. Although various technics have been proposed for testing qualitative data for unidimensionality, the technic of scale analysis, as developed by Guttman (16) for the field of attitude measurement, seems to offer the best method. One criticism which has been leveled against scale analysis, however, is that the initial selection of items is left to the intuition of the investigator. To avoid this criticism, Edwards and Kilpatrick (18) suggest that scale analysis be preceded by Thurstone scaling and item analysis in order to obtain a set of items which would have greater assurance of meeting the requirements of scale analysis. The study, herein reported, describes the successful application of the Edwards-Kilpatrick method to the development of a fatigue scale and the subsequent validation of this scale in two laboratory situations.

Received for publication on 1 May 1956.

DEVELOPMENT OF THE EXPERIMENTAL CHECKLIST

A fatigue scale to be useful in operational studies must not only meet the usual measurement criteria of reliability and validity, but also must be short, easy to fill out, easily understood, and simple to score. A type of rating scale which is growing in popularity and which best satisfies the requirements listed above is the checklist wherein the individual is required to make some response to each of a series of words, phrases, or descriptions (19).[1] The first research problem then was one of finding those words and phrases which would define a fatigue continuum and which would be understandable to a large, heterogeneous population.

Item selection

Fifteen airmen working in the department laboratories were asked to list words and phrases which might describe a fatigue continuum. The continuum, at this stage, was roughly defined as one extending from extreme tiredness on one end to extreme well-being on the other. The senior author also searched dictionaries and thesauri for appropriate items. Altogether approximately 500 items were collected. These were then screened by the above individuals against three criteria: (1) Would the item be familiar to individuals from all areas of the United States, or did the item have a regional flavor? (2) Was the vocabulary level of the item appropriate for available

[1] In this paper we may consider the checklist as a type of attitude scale wherein the individual is required to indicate his "attitude" toward his state of fatigue.

research populations? (3) Did the item fit the defined fatigue continuum, or could it connote other affective states such as anxiety, boredom, motivation, or morale?

One hundred fifty items survived this initial screening. Four psychologists of the department discussed each of these items, accepting it or rejecting it as belonging to a fatigue continuum.[2] Meanwhile, lists of the items were presented to 100 male basic Air Force trainees with the instructions to indicate any items which they did not recognize or understand. The above two procedures reduced the number of items to 92. The next problem, then, was to determine where each of these items belonged on the fatigue continuum. This, of course, was a problem of scaling.

Thurstone scaling

The 92 items, which at this stage constitute a sample of the universe of fatigue items, were typed separately on 1½ x 3½ inch cards. The problem was then outlined to twelve qualified research personnel who had agreed to serve as judges. Following a brief review of Thurstone's method of equal-appearing intervals (19), the investigator presented the cards to each judge, in turn, with instructions to sort them along a nine-interval continuum – interval 1 representing extreme well-being, interval 9 representing extreme fatigue. After every sorting, the investigator recorded the interval in which each item had been placed, then randomized the cards for presentation to the next judge.

The data available upon completion of the scaling process by all twelve judges provided for computation of scale (median) values and ambiguity (Q) values. Forty-eight items were rejected as being ambiguous. Thus, with 44 items available for further analysis, the first step in the Edwards-Kilpatrick procedure had been completed.

The experimental checklist

In developing the format for the experimental checklist, it was necessary to decide on some type of response system. Since scale analysis demands at least dichotomous response categories, the system used by Frazier (15) seemed ideal. This system offers the subject a choice for each item of one of three response categories: better-than, same-as, or worse-than. One advantage of this system which is particularly appropriate in the measurement of fatigue is that the subject is forced to pin-point himself on the fatigue continuum.

The experimental checklist, given the noncommittal name "Feeling-Tone Checklist," is shown in the appendix as figure 4. The items, it is to be noted, are randomly ordered. Separate checklist instructions were developed for use in the experiment which is to be described.

[2]The group agreed in rejecting such items as: fine, good, O.K., fair, swell, terrific, super, lousy, miserable, awful, and pretty low. These may belong to another continuum, possibly one which could be termed "elation-depression."

DEVELOPMENTAL STUDY

The developmental study was designed to provide data for item validity estimates and internal consistency item analyses. To test the items for validity it was necessary to find a suitable criterion; that is, a situation had to be created which would definitely produce fatigue. Valid items would, of course, discriminate significantly between a "nonfatigue" situation and the "fatigue" situation. The apparatus chosen to produce such a "fatigue" situation will now be described.

The USAF SAM Multidimensional Pursuit Test (CM 813 E) has been fully described elsewhere (20, 21). Test subjects are required to manipulate throttle, stick, and rudder controls so as to compensate for the apparently random movements of four instrument pointers from their null positions. When all four pointers are centered concurrently a timer cumulates an accuracy score in units of 0.01 minute.

Subjects performing on the test apparatus not only manifest task aversion both subjectively and objectively, but also complain of tiredness in specific body locations (20, 22). Decline in task proficiency (work decrement) is evidenced after about an hour's practice. Since the task simulates certain perceptual-motor components of the pilot's task, it is a particularly appropriate one for use in the present study if we are to extend our results to the problem of aircrew fatigue.

In the present study two copies of the test apparatus were used to test subjects in pairs. A common cycling device metered out alternate work trials (1-minute) and rest periods (15-second) for any desired span of time.

Subjects

The experimental sample consisted of 48 volunteer, experimentally naive, male basic Air Force trainees. Subjects were judged to have had adequate rest and to be otherwise fit for the task.

Procedure

At 0900 hours each testing day a qualified examiner read the checklist instructions to the subjects who then proceeded to fill out the experimental form. Immediately following this, they were instructed in the operation of the test apparatus, then received 40 trials of initial learning (0915-1005 hours) to establish a substantial level of skill. This was followed by a 10-minute interval during which they received two indoctrinations.

The first indoctrination, described as I_2 in another paper (20), related the important part played by the subject in helping the Air Force solve one of its research problems, and encouraged the subject to keep alert and do his best throughout the task to follow. The second indoctrination described a performance feedback device (peg-board) which was to provide the subject with knowledge of his standing at the end of each cycle (1 cycle = 8 trials, or about 10 minutes' time). A detailed description of this motivational treatment appears elsewhere (23). It was thought that the first indoctrination would induce a favorable task attitude, whereas knowledge of results would maintain (i.e., motivate) task interest and repress the onset of boredom. Such precautions, it was argued, were necessary if the checklist was to reflect fatigue only.

After the examiner had delivered the feedback indoctrination, the subjects were put back to work on the task for 4 hours (1015-1415 hours). At the conclusion of the test period, the subjects again filled out the experimental checklist.

Testing was conducted in an air-conditioned, well-illuminated room. Subjects could not observe one another's performance.

Results

The results of the analyses performed on the checklist data are shown in table I. As a matter of convenience the items were ordered continuum-wise. Each item's place on the continuum was judged by the investigator on the basis of the response category frequencies. Item validity was inferred from chi-square tests of significance. An item was accepted as valid when it could be shown to discriminate significantly between fatigued and nonfatigued criterion groups. The response category frequencies resulting from the first administration (A.M.) constituted the "nonfatigued" (or to be more explicit, "less-fatigued") criterion data; the response category frequencies from the second administration (P.M.) constituted the "fatigued" criterion data.

Internal consistency item analyses were performed on both A.M. and P.M. data. Checklists were scored using simple weights as follows: "Better than" response, 2; "Same as" response, 1; "Worse than" response, 0. Each set of data was then divided into high score (N = 18) and low score (N = 18) criterion groups. Chi square was then used to test the significance of difference between the marginal frequencies of the two criterion groups. The results shown in table I reveal a definite trend: significant items for the A.M. data are predominantly from the positive end of the continuum, while those for the P.M. data are predominantly from the negative end of the continuum.

Construction of equivolent-form checklists

Since only a handful of items could be rejected on the basis of poor validity and internal consistency (e.g., items 14, 15, 16, 30, 43, 44), the large number of "good" items remaining offered the possibility of constructing two checklists rather than one. The construction of equivalent-form checklists offers two advantages: (1) it provides a measure of reliability; (2) it adds versatility to one's research program. Since these checklists were to be subjected to scale analysis as part of the Edwards-Kilpatrick method, and as a typical Guttman scale consists of 10 to 12 items, a shortage of good items did not seem to be a problem. However, to find pairs of "equivalent" items from those available was a different matter. The chief criterion used in pairing items was whether the two items had similar response category frequencies. Items from intervals 1 through 4 for the A.M. data and from intervals 6 through 9 for the P.M. data with internal consistency probability levels of greater than .05 were not considered. Item 33 was eliminated as it proved to be ambiguous for several test subjects. Items 41 and 42 were chosen to "anchor" the checklists even though not valid in terms of the data; it was hypothesized that these items would prove valid under more fatiguing conditions. At this point it was not possible to select equivalent items for scale intervals 3 and 7, and therefore two items had to serve "double duty" on both forms. No items were used from the "neutral," or middle, zone (interval 5) in accordance with Edwards and Kilpatrick's caution (24). The items for the two checklists, designated as Form A and Form B, are shown, randomly ordered, in figure 1.

In the appendix, figure 5, is shown the checklist format which was used in the validation study to be described.

TABLE I

Results of analysis of the experimental checklist data

Item	Marginal frequencies						Probability level		
	A. M. data			P. M. data			Item validity	Internal consistency	
	Better than	Same as	Worse than	Better than	Same as	Worse than		A. M. data	P. M. data
Scale Interval 1									
1. like I'm bursting with energy	1	15	32	0	1	47	.001	.001	.70
2. I never felt fresher	2	15	31	0	2	46	.001	.001	.50
3. I never felt more peppy	0	19	29	0	1	47	.001	.001	.70
4. extremely lively	1	19	28	0	2	46	.001	.001	.70
5. extremely fresh	0	21	27	0	2	46	.001	.001	.50
6. extremely peppy	2	19	27	0	2	46	.001	.001	.50
7. extremely refreshed	1	26	21	0	2	46	.001	.001	.50
Scale Interval 2									
8. very peppy	0	23	25	0	1	47	.001	.001	.70
9. very fresh	2	25	21	0	2	46	.001	.001	.50
10. very lively	2	25	21	0	4	44	.001	.01	.10
11. very rested	2	26	20	1	3	44	.001	.01	.20
12. very refreshed	1	30	17	0	1	47	.001	.01	.70
Scale Interval 3									
13. quite fresh	2	35	11	0	2	46	.001	.05	.50
14. quite lively	2	40	6	0	6	42	.001	.30	.02
15. quite rested	2	41	5	1	2	45	.001	.30	.20
16. pretty fresh	7	40	1	0	7	41	.001	.10	.01
Scale Interval 4									
17. a little peppier than usual	4	23	21	0	2	46	.001	.001	.50
18. a little fresher than usual	4	33	11	0	1	47	.001	.001	.70
19. somewhat peppy	8	37	3	1	14	33	.001	.05	.001
20. somewhat fresh	12	36	0	0	10	38	.001	.001	.001
21. somewhat refreshed	13	35	0	0	4	44	.001	.02	.10
22. a little peppy	13	35	0	1	17	30	.001	.001	.001
23. a little fresh	15	33	0	1	12	35	.001	.001	.001
Scale Interval 5									
24. no peppier than usual	16	31	1	2	8	38	.001	.001	.20
25. no fresher than usual	18	29	1	0	10	38	.001	.001	.05
26. not too tired, not too fresh	19	29	0	1	23	24	.001	.001	.001
Scale Interval 6									
27. a wee bit tired	24	24	0	3	16	29	.001	.001	.001
28. slightly tired	30	18	0	2	25	21	.001	.01	.001
29. a little tired	30	18	0	3	28	17	.001	.01	.001
30. somewhat tired	29	18	1	1	34	13	.001	.001	.20
31. slightly pooped	33	15	0	3	32	13	.001	.05	.05
32. a little pooped	39	9	0	4	32	12	.001	.30	.05
Scale Interval 7									
33. pretty tired	47	1	0	8	34	6	.001	.98	.001
34. fairly well pooped	44	3	1	9	35	4	.001	.70	.001
Scale Interval 8									
35. awfully tired	47	1	0	15	32	1	.001	.98	.001
36. petered out	47	1	0	17	30	1	.001	1.00	.001
37. very tired	48	0	0	22	25	1	.001	1.00	.001
38. tuckered out	48	0	0	22	26	0	.001	1.00	.001
Scale Interval 9									
39. weary to the bone	47	1	0	30	14	4	.001	.98	.01
40. extremely tired	47	1	0	33	13	2	.001	.98	.01
41. dead tired	48	0	0	36	11	1	.30	1.00	.001
42. ready to drop	48	0	0	38	8	2	.50	1.00	.01
43. completely exhausted	48	0	0	40	5	3	.70	.70	.20
44. knocked out	47	0	1	43	5	0	.70	.98	.50

VALIDATION STUDY I

Task

The task used in this study was the Multidimensional Pursuit Test which has been described above.

Subjects

The subjects were, as before, basic Air Force trainees, rested and fit for the task.

Procedure

The experimental subjects were taken into the test room at 0900 hours, read the checklist instructions, then asked to fill out the Feeling-Tone Checklist, Form A (this data will hereafter be referred to as "1A"). The subjects then received 40 trials of initial learning (0915-1005 hours) on the test apparatus after which they filled out the Feeling-Tone Checklist, Form B (hereafter, data 2B). This was followed by the I_2 and feedback indoctrinations previously described. Subjects then returned to their task for 3 hours (1015-1315 hours) at the conclusion of which they were given a 4-minute "rest period." During this period the subjects remained seated at their apparatus and filled out Form A of the checklist (data 3A). Subjects then were tested for an additional half hour (1319-1349 hours). At the conclusion of the testing program the subjects received first one form of the checklist and then the other to fill out (4A and 4B: assigned at random in a counterbalanced order).

Subjects were tested in pairs until a population of 100 was obtained. Experimental conditions were otherwise identical with those of the previously described experiment.

Concurrently with the testing of each pair of experimental subjects, pairs of control subjects were also "tested" in an adjacent room. These subjects received the same schedule of checklists as did the experimental subjects. A separate indoctrination given at the start of testing (0910 hours) told the subjects, in essence, that they were jet pilots on the alert and should be ready to "man their planes" on a moment's notice. It is believed that this indoctrination was successful in keeping the subjects alert during their 4½-hour test period. When not engaged in filling out checklists, the control subjects were allowed to read magazines, converse, write letters, and smoke.

Form A	Item No.	Form B
slightly tired	1	a little tired
like I'm bursting with energy	2	I never felt fresher
extremely tired	3	weary to the bone
quite fresh	4	quite fresh
slightly pooped	5	a little pooped
extremely peppy	6	extremely lively
somewhat fresh	7	somewhat refreshed
petered out	8	awfully tired
very refreshed	9	very rested
ready to drop	10	dead tired
fairly well pooped	11	fairly well pooped
very lively	12	very fresh
very tired	13	tuckered out

FIGURE 1
Checklist equivalent forms

Results

The procedure just described provided 10 sets of checklist data: 1A, 2B, 3A, 4A, 4B for both experimental and control subjects. The several analyses performed on these data will now be described.

First of all, the checklists were scored using the simple 2, 1, 0 weights described above. Product-moment correlations were then computed between subject scores of data 4A and 4B for both experimental and control groups. The resulting correlations, which are estimates of the equivalent-form reliability, were .92 for the experimental group and .95 for the control group.

The determination of Form A checklist validity was effected by a comparison of the feeling-tone level of the experimental and control groups at the first, third, and fourth checklist administrations. A repeated measurements analysis was made within the split-plot design of Groups (2) x Administrations (3). Results are shown in table II where one should note the significant A x G interaction which points to a difference in slope for the group mean curves. This may be interpreted as meaning there is a

TABLE II
Analysis of variance of Form A checklist data

Source of variance	df	Mean square	F	P
Groups	1	1,744.22	62.81	.001
Error (a)	198	27.77		
Administrations	2	2,011.17	269.23	.001
A x G	2	386.40	51.73	.001
Error (b)	396	7.47		
Total	599			

greater decrease in feeling-tone over time for the experimental as compared with the control group. This finding is all the more noteworthy when one considers the fact that the feeling-tone level of the control subjects showed a significant decline between the first and third administrations as demonstrated by a t-test ($t = 3.80$; $P < .001$) between the subclass means (see table IIIA). Thus the ability of this checklist to reflect a significantly greater decline in feeling-tone for the experimental group, when both experimental and control groups became significantly "tired" in terms of checklist data, is more than adequate proof of its validity.

The determination of Form B checklist validity was effected in a similar manner. Here the comparison was between the feeling-tone level of the experimental and control groups at the second and fourth administrations. Results of the repeated measurements analysis made within the split-plot design of Groups (2) x Administrations (2) are shown in table IV. Once again the significant A x G interaction points to a greater decrease in feeling-tone over time for the experimental as compared with the control group. Subclass means for the Form B data are shown in table IIIB. The difference between the control and experimental groups at the second administration is significant ($t = 3.98$; $P < .001$); and, as was the case with the Form A data, the feeling-tone level of the control subjects showed a significant decline with time ($t = 4.51$; $P < .001$—between data 2B and 4B).

Figure 2 plots the mean level of subjective fatigue for both control and experimental groups throughout the 4½-hour task. The two curves merely corroborate the statistical evidence for validity of the two checklist forms.

Thus having analyzed the data for reliability and validity, the next step was to test it against a third criterion, unidimensionality. This, as mentioned previously, is accomplished by scale analysis. The method used in this study was a modification of Guttman's Cornell technic as described by Niven (25) plus the improvement suggested by Menzel (26). The procedure of analysis was as follows: A perfect scale was derived on the basis of *all* the checklist data. Separate scale analyses were then made on data 2B, 3A, 4A, and 4B for both experimental and control groups plus one analysis on the 1A data combined for both groups. The merging of the

TABLE III

A. Subclass means for Form A checklist administrations

Group	Administrations			Rows
	No. 1	No. 3	No. 4	
Control	17.77	14.91	14.37	15.68
Experimental	17.51	9.39	9.92	12.27
Columns	17.64	12.15	12.14	13.98

B. Subclass means for Form B checklist administrations

Group	Administrations		Rows
	No. 2	No. 4	
Control	17.08	14.15	15.61
Experimental	14.49	9.94	12.21
Columns	15.78	12.04	13.91

TABLE IV

Analysis of variance of Form B checklist data

Source of variance	df	Mean square	F	P
Groups	1	1,156.00	53.79	.001
Error (a)	198	21.49		
Administrations	1	1,398.76	200.10	.001
A x G	1	65.61	9.38	.005
Error (b)	198	6.99		
Total	399			

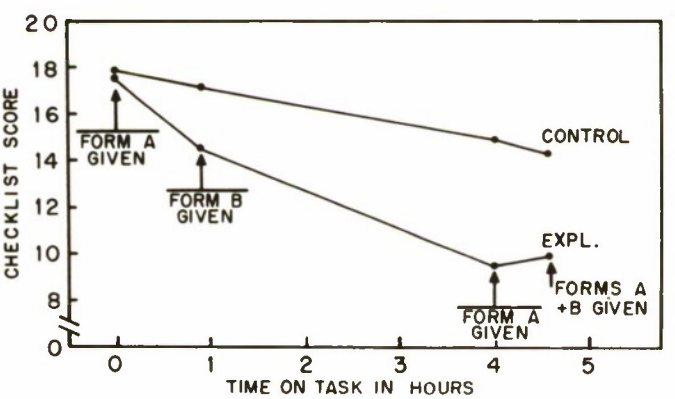

FIGURE 2

The course of subjective fatigue under 4½-hour task and no-task conditions.

control and experimental data was justified since the "experimental" subjects had yet to be subjected to different experimental conditions than the controls. The individual papers within each of these nine sets of data, having already been scored for previous analyses, were then ordered from high to low. A scatterplot was then made with item responses being tallied in columns headed by the item response categories and on the same line in which the individual's total score was recorded. An error was recorded whenever a response fell outside the perfect scale pattern. Guttman's coefficient of reproducibility was then determined by computing the percent of consistent responses. Menzel's coefficient of scalability, which is essentially a measure of efficiency (it measures the amount of reproducibility which could not be attained in the absence of scalability), was next obtained. Results of this scale analysis (first approximation) are shown in table V. A coefficient of reproducibility of .90 is considered acceptable. Although no specific level of acceptance is presently recognized for the coefficient of scalability, Menzel suggests that it may be somewhere between .60 and .65.

With respect to individual items, no item had a reproducibility coefficient low enough to consider dropping it from the checklist. Two pairs of equivalent items, however, were noted to have item response frequencies similar to other items and therefore were performing superfluous jobs. By eliminating items 2 and 13 from Form A and items 2 and 8 from Form B it was possible to maintain equivalency while reducing the number of items in the checklists to 11 each. To obtain reproducibility and scalability coefficients for the 11-item checklist "forms," the data were rescored and then evaluated by the procedures of scale analysis previously described. Inspection of the results indicated that one item, item 1 (both forms), was responsible for a considerable proportion of the errors of reproducibility. Consequently this item was eliminated. The data were then rescored on the basis of the 10 remaining items and re-evaluated by scale analysis. Results of both second and third approximations are shown in table IV. Discussion of these results will follow the second checklist validation study about to be described.

VALIDATION STUDY II

This study, although not a validation study in the true sense, was designed to test the sensitivity of the checklist to drug-induced changes in affective state. The need for measuring the subjective response in drug research was outlined by Beecher in 1952 (27). Since then the inclusion of subjective questionnaires as part of research programs has provided greater insight into the results of studies by Lasagna et al. (28) and Nowlis et al. (29, 30).

In the present study two drugs which are known to induce dissimilar task dispositions (23) were

TABLE V

Results of scale analyses

Data	N	Reproducibility (%)			Scalability (%)		
		First approximation	Second approximation	Third approximation	First approximation	Second approximation	Third approximation
1A*	200	91.38	91.09	93.20	65.43	64.23	70.37
3A-Expl	100	89.54	89.45	90.80	61.80	61.20	63.05
4A-Expl	100	89.62	89.45	90.40	57.81	56.39	58.08
3A-Cont	100	90.00	89.45	90.40	64.38	64.42	66.55
4A-Cont	100	91.00	90.55	92.40	62.01	63.25	70.31
2B-Expl	100	87.85	86.09	89.40	55.99	52.04	63.82
4B-Expl	100	89.92	89.18	90.00	57.33	54.75	55.56
2B-Cont	100	90.31	89.91	91.00	60.87	60.50	61.37
4B-Cont	100	89.08	88.45	89.60	55.76	57.09	61.34

*See text for code.

used. These were the analeptic, dexedrine (5 mg. dextro-amphetamine sulfate), and the depressant, benadryl-hyoscine (0.65 mg. hyoscine hydrobromide mixed with 50 mg. diphenhydramine hydrochloride). Lactose placebo was used in the study as a control. All drugs were administered in No. 1 pink capsules.

Air Force basic trainees, having been screened for contraindications to drug administration, served as subjects. The task was similar to that of the control subjects as described in the previous section; in other words, the subjects were not required to perform in any experimental situation but were allowed to converse, read magazines, and write letters. Subjects were observed continuously for a period of 4½ hours. The test area was air-conditioned and well illuminated.

The procedure was as follows: Each day at 0930 hours the subjects, in groups of three to six, heard an examiner read the checklist instructions, then proceeded to fill out the Form A checklist. The examiner then described the task and presented the subjects with their capsules according to a previously randomized schedule. Subjects were given the chance to refuse to participate in the experiment if they so desired. The "task" proceeded with the Form A checklist being administered at 1½-hour intervals (1100, 1230, and 1400 hours). Testing was conducted until a population of 120 subjects (40 per drug group) was attained.

To appraise the effect of drugs on feeling-tone over the 4½-hour period the following was accomplished. Individual checklist scores from the second, third, and fourth administrations were adjusted for regression upon the initial administration. A repeated measurements analysis was then made within the split-plot design of Groups (3) x Administrations (3). Results are shown in table VI. Although the significant A x T interaction indicates that, over time, the groups did not respond in a similar fashion, the picture is clarified by testing for significant differences (by t-tests) between the subclass means (table VII). For the second and third administrations both dexedrine and placebo subgroups are significantly different from the benadryl-hyoscine subgroup; for the fourth administration there were no significant differences between subgroups. Supporting the statistical findings are the feeling-tone curves, shown in figure 3, which were plotted from the means of the checklist scores. The statistical analyses and graphic presentation taken together demonstrate the sensitivity of the checklist to drug-induced changes in affective state.

The 120 checklists from the first administration provide additional data with which one can check for unidimensionality. The data were scored not only for the complete 13-item form, but also for the

TABLE VI

Analysis of adjusted variance of checklist data

Source of variance	df	Mean square	F	P
Treatments	2	171.62	5.76	.02
Error (a)	117	29.82		
Administrations	2	45.46	4.24	.02
A x T	4	134.88	12.58	.001
Error (b)	234	10.72		
Total	359			

TABLE VII

Subclass means of adjusted checklist scores

Treatment	Administrations			Rows
	No. 2	No. 3	No. 4	
Placebo	16.7	14.9	14.5	15.4
Dexedrine	16.3	16.4	15.4	16.0
Benadryl-hyoscine	13.2	11.5	13.3	12.7
Columns	15.4	14.3	14.4	14.4

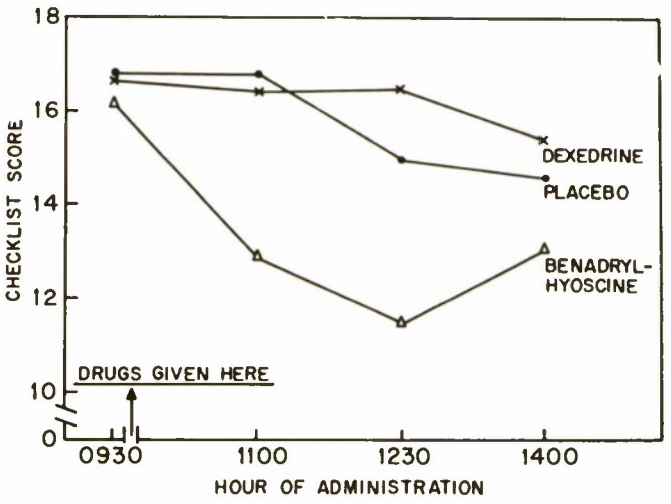

FIGURE 3

The effects of pharmacologic treatments upon the course of subjective fatigue.

11- and 10-item "forms" described above. Results of the scale analyses performed thereon gave reproducibility coefficients of 92.12, 92.20, and 92.75 and scalability coefficients of 68.70, 70.90, and 70.51 for the first, second, and third approximations, respectively.

DISCUSSION

That the checklists are equivalent and that their items, both individually and collectively, are valid has been demonstrated. It is, however, somewhat difficult to give an unequivocal "yes" or "no" answer to the question of whether the checklists are unidimensional. This is a problem commonly faced by an investigator employing scale analysis since many of the criteria are subjective, and it is the investigator alone who must provide the answer. The coefficients of reproducibility exceed or closely approach the .90 acceptance level. Not taken into account by Guttman's coefficient is human error – that is, obviously misplaced checkmarks, of which there was a considerable amount for the population used in the studies. On the other hand, the extreme response frequencies of items such as "I never felt fresher" and "dead tired" result in what is termed *artificial reproducibility*. Yet, it is argued that in the case of constructing a fatigue checklist, admittedly one peculiar to the field of attitude measurement, items from both *ends* of the continuum are required. Under nonfatigue conditions "extremely peppy" is functioning at its best whereas "ready to drop" is not functioning at all; yet, under extreme fatigue conditions the reverse is true. The item response frequencies for each checklist, obtained under the various conditions described, adequately cover the continuum and should, *in toto*, reflect the subjective state for any fatigue-research situation conceivable.

Final evaluation of the checklists must wait until their usefulness can be demonstrated in laboratory and operational studies of fatigue: its manifestations, etiology, mitigation, and consequences. Particular note should be made of the inability of past studies to explain those relationships found between subjective fatigue and work decrement. For example, it is quite possible for a subject to improve his performance on a simple task lasting for 2 to 4 hours while his feelings of tiredness increase. The flight surgeon, however, would be interested in comparing changes in affective state with aircrew performance on prolonged flights, and, if a high correlation exists between the two measures, determining a level of subjective fatigue where performance can be expected to be critically affected. Studies aimed at discovering the importance of personal equipment, cockpit design, mission stresses, analeptic drugs, sleep characteristics, and diet as related to the affective state would be of concomitant interest.

The checklists developed in the study described are easily understood and take but a few seconds to fill out. Such qualities are highly desirable if the checklists are to be filled out by aircrew personnel performing their duties in flight. A suggested set of instructions and 10-item checklist (Form A) for future research appears in the appendix, figures 6 and 7.

SUMMARY

In a developmental study designed to screen items for a subjective fatigue scale, 48 subjects were tested on a complex, fatiguing perceptual-motor task for a period of 5 hours. A 44-item experimental checklist was given before and after psychomotor testing. Two 13-item, equivalent-form checklists were constructed from items which demonstrated both validity and internal consistency. In a validation study 100 subjects were tested on the perceptual-motor task for 4½ hours. Checklist administrations were scheduled to provide data for testing unidimensionality (by Guttman scale analysis), for an estimate of validity, and for estimating reliability. One hundred control subjects who were allowed to relax, read magazines, and write letters, received the same schedule of checklists. Results were as follows: (a) Equivalent-form reliability was .92 and .95 for the experimental and control groups, respectively; (b) although both an experimental group and a control group became significantly "tired" in terms of checklist data, such data were able to reflect a significantly greater decline in feeling-tone for the experimental group; (c) the data seemed to meet adequately the requirements of scale analysis as to unidimensionality.

In a related study 120 subjects, after being administered the checklist, were equally and randomly assigned to one of three drug treatments (analeptic, depressant, or placebo). Subjects performed no task but were administered checklists at 90-minute intervals over the next 4½ hours. The results clearly demonstrated the ability of

the checklist to reflect the expected differences in affective state.

A 10-item checklist form was described and suggested for use in future research.

The authors are indebted to A/1C Jack F. Knoche, A/1C Richard H. Johnson, and A/1C Claire E. Steele for assistance in the collection of experimental data, and to S/Sgt. Samuel S. Mann, Jr., and A/1C Robert T. Stanton for statistical analyses. Captain Edwin W. Moore provided invaluable assistance and criticism on matters of experimental design and analysis.

REFERENCES

1. Bartley, S. H., and E. Chute. Fatigue and impairment in man. New York: McGraw-Hill, 1947.
2. Ryan T. A. Work and effort. New York: Ronald Press, 1947.
3. Floyd, W. F., and A. T. Welford (Eds.). Symposium on fatigue. London: H. K. Lewis and Co., Ltd., 1953.
4. Ryan, T. A., and P. C. Smith. Principles of industrial psychology. New York: Ronald Press, 1954.
5. Thorndike, R. L. The human factor in accidents with special reference to aircraft accidents. School of Aviation Medicine, USAF, Project No. 21-30-001, Report No. 1, February 1951.
6. Gugenheim, C. Étude experimentale des aspects objectifs et subjectifs de la fatigue pendant un travail monotone. I. Recherche effectuée sur un group d'etudiants. Travail hum. 16: 219-240 (1953).
7. Gross, I. H., and S. H. Bartley. Fatigue in house care. J. Appl. Psychol. 35:205-208 (1951).
8. Griffith, J. W., W. A. Kerr, T. B. Mayo, and J. R. Topal. Changes in subjective fatigue and readiness to work during the eight-hour shift. J. Appl. Psychol. 34:163-166 (1950).
9. Cooper, B. J. The relationship of psychological fatigue to work performed. Unpublished M. S. thesis, Western Reserve University, Cleveland, Ohio, 1953.
10. Muscio, B. Feeling-tone in industry. Brit. J. Psychol. 12:150-162 (1921).
11. Poffenberger, A. T. Effects of continuous work upon output and feelings. J. Appl. Psychol. 12:459-467 (1928).
12. Barmack, J. E. The time of administration and some effects of 2 grs. of alkaloid caffeine. J. Exper. Psychol. 27:690-698 (1940).
13. Middleton, W. C., P. J. Fay, W. A. Kerr, and F. Amft. The effect of music on feelings of restfulness-tiredness and pleasantness-unpleasantness. J. Psychol. 17:299-318 (1944).
14. Foltz, E. E., F. T. Jung, and L. E. Cisler. The effect of some internal factors on human work output and recovery. Amer. J. Physiol. 141: 641-646 (1944).
15. Frazier, G. The development of a scale of subjective feelings of fatigue. Unpublished M. S. thesis, Purdue University, Lafayette, Ind., 1952.
16. Stouffer, S. A., L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen. Measurement and prediction. Princeton: Princeton University Press, 1950.
17. McNelly, G. W. The development and laboratory-validation of a subjective fatigue scale. Dissertation Abstr. 14:2135 (1954).
18. Edwards, A. L., and F. P. Kilpatrick. A technique for the construction of attitude scales. J. Appl. Psychol. 32:374-384 (1948).
19. Guilford, J. P. Psychometric methods (2d ed.) New York: McGraw-Hill, 1954.
20. Payne, R. B., and G. T. Hauty. The effects of experimentally induced attitudes upon task proficiency. J. Exper. Psychol. 47:267-273 (1954).
21. Melton, A. W. (Ed.). Apparatus tests. Report No. 4 (restricted supplement). AAF Avn. Psychol. Program Rsch. Reports. Washington: U. S. Government Printing Office, 1947.
22. Pearson, R. G., and R. O. Bauer. The effects of morphine-nalorphine mixtures on psychomotor performance. School of Aviation Medicine, USAF, Report No. 55-137, June 1956.
23. Payne, R. B., and G. T. Hauty. The influence of drugs, motivation, and job design upon work decrement. School of Aviation Medicine, USAF, Report No. 55-16, Jan. 1955.
24. Edwards, A. L., and F. P. Kilpatrick. Scale analysis and the measurement of social attitudes. Psychometrika 13:99-114 (1948).
25. Niven, J. R. A comparison of two attitude scaling techniques. Educ. Psychol. Measmt. 13:65-76 (1953).

26. Menzel, H. A new coefficient for scalogram analysis. Publ. Opin. Quart. 17:268-280 (1953).

27. Beecher, H. K. Experimental pharmacology and measurement of the subjective response. Science 116:157-162 (1952).

28. Lasagna, L., J. M. von Felsinger, and H. K. Beecher. Drug-induced mood changes in man. 1. Observations on healthy subjects, chronically ill patients, and "postaddicts." J.A.M.A. 157:1006-1020 (1955).

29. Nowlis, V. The development and modification of motivational systems in personality. *In* Brown, J. S., et al. Current trends and research in motivation: A symposium. Lincoln, Nebr.: University of Nebraska Press, 1953.

30. Nowlis, H. H., V. Nowlis, A. H. Riesen, and G. R. Wendt. Chemical influences on behavior. III. The effects of dramamine and scopolamine on emotional and social behavior with comparison data on the effects of other drugs. U. S. Navy, Office of Naval Research Project NR 144-060; 54000, 21 Mar. 1953.

APPENDIX

FEELING-TONE CHECKLIST

Name_____ AFSN_____ Date_____

T.SQ._____ FLT._____ Home State_____ Age_____

INSTRUCTIONS: Consider the statement in question, and determine in your own mind whether you feel, at this instant, "Better than," the "Same as," or "Worse than" the feeling described by the statement. Then place a check mark in the appropriate box.

No.	Better Than	Same as	Worse Than	Statement
1.	( )	( )	( )	quite lively
2.	( )	( )	( )	extremely refreshed
3.	( )	( )	( )	awfully tired
4.	( )	( )	( )	I never felt more peppy
5.	( )	( )	( )	dead tired
6.	( )	( )	( )	very refreshed
7.	( )	( )	( )	a wee bit tired
8.	( )	( )	( )	petered out
9.	( )	( )	( )	not too tired, not too fresh
10.	( )	( )	( )	a little fresher than usual
11.	( )	( )	( )	very peppy
12.	( )	( )	( )	ready to drop
13.	( )	( )	( )	like I'm bursting with energy
14.	( )	( )	( )	pretty tired
15.	( )	( )	( )	a little fresh
16.	( )	( )	( )	extremely lively
17.	( )	( )	( )	no fresher than usual
18.	( )	( )	( )	weary to the bone
19.	( )	( )	( )	somewhat tired
20.	( )	( )	( )	quite fresh
21.	( )	( )	( )	completely exhausted
22.	( )	( )	( )	no peppier than usual
23.	( )	( )	( )	extremely fresh
24.	( )	( )	( )	somewhat peppy
25.	( )	( )	( )	very tired
26.	( )	( )	( )	quite rested
27.	( )	( )	( )	very fresh
28.	( )	( )	( )	fairly well pooped
29.	( )	( )	( )	knocked out
30.	( )	( )	( )	I never felt fresher
31.	( )	( )	( )	very lively
32.	( )	( )	( )	extremely tired
33.	( )	( )	( )	a little peppy
34.	( )	( )	( )	tuckered out
35.	( )	( )	( )	a little pooped
36.	( )	( )	( )	extremely peppy
37.	( )	( )	( )	a little tired
38.	( )	( )	( )	pretty fresh
39.	( )	( )	( )	a little peppier than usual
40.	( )	( )	( )	slightly tired
41.	( )	( )	( )	somewhat refreshed
42.	( )	( )	( )	very rested
43.	( )	( )	( )	slightly pooped
44.	( )	( )	( )	somewhat fresh

FIGURE 4

FEELING TONE CHECKLIST, FORM A

Name_____ AFSN_____ Date_____

T.SQ._____ FLT._____ Home State_____ Age_____

INSTRUCTIONS:

People feel different at various times for various reasons. Some arise after a night's rest feeling "quite rested" while others may feel "a little tired." A hard day's work or a vigorous workout at the gym may make you feel "fairly well pooped"; yet, a shower, a cup of coffee, or merely a few minutes relaxing in a comfortable chair may make you feel "very refreshed."

We would like to find out how you feel right now. Below you will see 13 statements which describe different degrees of freshness or peppiness and tiredness. For each statement you will have to determine in your own mind whether you feel at this instant (1) "Better than," (2) the "Same as," or (3) "Worse than" the feeling described by that statement. Having done this you will then place an "X" in the appropriate box.

Consider the following example:

No.	Better than	Same as	Worse than	Statement
0.	( )	( )	( )	somewhat tired

If right now you felt "somewhat tired" you would place an "X" in the box marked "Same as." If, however, you felt fresh or full of pep you would check the box marked "Better than" because you would be feeling better than "somewhat tired." On the other hand, if you felt exhausted you would place an "X" in the box marked "Worse than."

Take each statement in order; do not skip around from one to another. Read each statement carefully so that you understand what it means. It may help you to understand some statements if you insert the words "I feel" or "I am" before the statement.

This is not a test. You have all the time you need.

No.	Better than	Same as	Worse than	Statement
1.	( )	( )	( )	slightly tired
2.	( )	( )	( )	like I'm bursting with energy
3.	( )	( )	( )	extremely tired
4.	( )	( )	( )	quite fresh
5.	( )	( )	( )	slightly pooped
6.	( )	( )	( )	extremely peppy
7.	( )	( )	( )	somewhat fresh
8.	( )	( )	( )	petered out
9.	( )	( )	( )	very refreshed
10.	( )	( )	( )	ready to drop
11.	( )	( )	( )	fairly well pooped
12.	( )	( )	( )	very lively
13.	( )	( )	( )	very tired

Have you checked each statement?

FIGURE 5

Instructions for Operational Use

INSTRUCTIONS: The statements to follow are to help you decide how you feel at this time -- not yesterday, not an hour ago -- but right now. For each statement you must determine whether you feel (1) "Better than," (2) "Same as," or (3) "Worse than" the feeling described by that statement.

As an example, take a person who feels a little tired. He might respond to the following items as follows:

	Better than	Same as	Worse than	Statement
a)	( )	( )	(X)	extremely fresh
b)	( )	(X)	( )	slightly tired
c)	(X)	( )	( )	completely exhausted

In other words, this person feels worse than "extremely fresh," about the same as "slightly tired," but, on the other hand, better than "completely exhausted."

Now, answer each of the following statements as follows:

If you feel better than the statement, place an "X" in the "better than" column.

If you feel about the same as the statement, place an "X" in the "same as" column.

If you feel worse than the statement, place an "X" in the "worse than" column.

Remember, answer each question with regard to how you feel at this instant.

FIGURE 6

FEELING TONE CHECKLIST SF-1

No.	Better than	Same as	Worse than	Statement
1.	( )	( )	( )	very lively
2.	( )	( )	( )	extremely tired
3.	( )	( )	( )	quite fresh
4.	( )	( )	( )	slightly pooped
5.	( )	( )	( )	extremely peppy
6.	( )	( )	( )	somewhat fresh
7.	( )	( )	( )	petered out
8.	( )	( )	( )	very refreshed
9.	( )	( )	( )	fairly well pooped
10.	( )	( )	( )	ready to drop

FIGURE 7

DEPARTMENT OF THE AIR FORCE
SCHOOL OF AVIATION MEDICINE, USAF
RANDOLPH AFB, TEXAS
OFFICIAL BUSINESS

BULK RATE
U. S. POSTAGE
PAID
RANDOLPH AFB, TEXAS
PERMIT NO. 11

[illegible] TECH INFO CENTER
[illegible]
[illegible]
[illegible] BUILDING
DAYTON 2, OHIO